2nd January 2020

Categorizing Humanities clusters

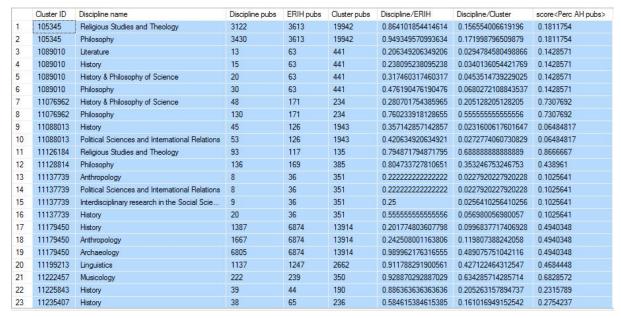
Using their ERIH PLUS disciplines

Understanding the disciplines

	ID	Discipline name	ah	SS
1	1	Archaeology	1	0
2	3	Musicology	1	0
3	5	Classical Studies	1	0
4	6	Literature	1	0
5	8	Film and Theatre Studies	1	0
6	10	Philosophy	1	0
7	11	Media Studies and Communication	1	1
8	13	Linguistics	1	1
9	16	Library and Information Science	1	0
10	18	Religious Studies and Theology	1	0
11	21	Cultural Studies	1	0
12	22	History	1	0
13	23	Interdisciplinary research in the Humanities	1	0
14	25	History & Philosophy of Science	1	0
15	26	Art and Art History	1	0
16	27	Interdisciplinary research in the Social Sciences	0	1
17	28	Gender Studies	0	1
18	29	Political Sciences and International Relations	0	1
19	24	Economics	0	1
20	0	Demography	0	1
21	19	Anthropology	0	1
22	20	Sociology	0	1
23	17	Human Geography and Urban Studies	0	1
24	14	Science and Technology Studies	0	1
25	15	Psychology	0	1
26	12	Social Statistics and Informatics	0	1
27	9	Pedagogical & Educational Research	0	1
28	7	Law	0	1
29	4	Business and Management	0	1
30	2	Environmental Studies	0	1

crossref

Only from main AH journals (equal amount or more AH disciplines than SS) For each cluster retrieve the ERIH PLUS disciplines

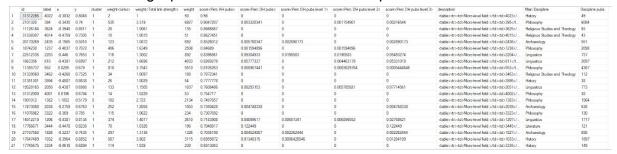


which ratio is more significant to display?

The number of pubs for discipline/number of total AH pubs from ERIH or over the total number of publications within the Cluster?

Solution with only main disciplines

N.B. The discipline is based on journals and a single journal can have multiple disciplines. This means that a single pub can be treated in more disciplines.



Working on displaying the graph.

- Gephi: Too much data. It crashes.
- VosViewer: Can't manage other kinds of colors
- Cytoscape: Layout in progress

Digital Humanities and Humanities

	id	label	x	У	cluster	weight < Links >	weight < Total link strength>	weight	score (Perc AH pubs)	score (Perc DH pubs)	score < Perc DH pubs level 1>	score < Perc DH pubs level 2>	description	Main Discipline	Discipline pubs
1	1880760	2409	-0.2336	0.7465	1	307	1.3318	718	0.4164345	0.6657382	0.270195	0.3955432	dable>dr>dd>Micro-level field:dd>2410 </td <td>Linguistics</td> <td>263</td>	Linguistics	263
2	1866507	3230	-0.3385	0.7744	1	82	1.0314	268	0.608209	0.5410448	0.4962687	0.04477612	dable>dr>dd>Micro-level field:dd>3231 </td <td>Linguistics</td> <td>111</td>	Linguistics	111
3	19513346	1004	0.2808	0.5548	5	613	2.6757	3219	0.3603604	0.3470022	0.01988195	0.3271202	dable>dr>dd>Micro-level field:dd>1005 </td <td>Linguistics</td> <td>1136</td>	Linguistics	1136
4	1704728	3646	0.5613	0.4038	5	90	1.0439	129	0	0.07751938	0	0.07751938	dable>dr>dd>Micro-level field:dd>3647 </td <td>NULL</td> <td>NULL</td>	NULL	NULL
5	1430295	1138	0.1912	0.5972	5	846	2.5371	2791	0.1640989	0.05589395	0.000716589	0.05517736	dable>dr>dd>Micro-level field:dd>1139 </td <td>Philosophy</td> <td>358</td>	Philosophy	358
6	254406	637	-0.2223	0.7429	1	889	2.43	4908	0.03198859	0.05093725	0.003463733	0.04747351	dable>dr>dd>Micro-level field:dd>638 <td>NULL</td> <td>NULL</td>	NULL	NULL
7	31315543	3607	-0.1544	0.5924	1	163	1.0493	142	0.09859155	0.04225352	0	0.04225352	dable>dr>dd>Micro-level field:dd>3608 </td <td>History</td> <td>3</td>	History	3
8	31315543	3607	-0.1544	0.5924	1	163	1.0493	142	0.09859155	0.04225352	0	0.04225352	dable>dr>dd>Micro-level field:dd>3608 </td <td>Musicology</td> <td>3</td>	Musicology	3
9	22512336	2203	-0.446	0.7653	1	116	1.3002	892	0.8396861	0.03811659	0.0190583	0.0190583	dable>dr>dd>Micro-level field:dd>2204 </td <td>Linguistics</td> <td>737</td>	Linguistics	737
10	1663817	2264	0.5677	0.3982	5	354	1.4092	844	0.01658768	0.03436019	0.02014218	0.01421801	dable>dr>dd>Micro-level field:dd>2265 </td <td>NULL</td> <td>NULL</td>	NULL	NULL
11	1874100	1255	-0.4046	0.705	1	809	2.3245	2459	0.4221228	0.03416023	0	0.03416023	dable>dr>dd>Micro-level field:d>1256 </td <td>Philosophy</td> <td>954</td>	Philosophy	954

The top 10 main DH clusters: (in crossref the DH clusters are related to Linguistic)

ID	weight	Perc AH pubs	Perc DH pubs	Main Discipline	Discipline pubs	Most frequent terms
		Pulle	Police	элоорино	Pane	<pre>authorship; analysis;</pre>
						attribution; using;
						language; text;
						study;
						classification;
2409	718	0.4164345	0.6657382	Linguistics	263	based; approach;
						english; digital;
						tei; early; analysis;
						case; texts; text;
3230	268	0.608209	0.5410448	Linguistics	111	edition; scholarly;
						analysis; sentiment;
						using; translation;
						language; machine;
						semantic; text;
1004	3219	0.3603604	0.3470022	Linguistics	1136	based; approach;
3646	129	0	0.07751938	NULL	NULL	NULL
1138	2791	0.1640989	0.05589395	Philosophy	358	NULL
637	4908	0.03198859	0.05093725	NULL	NULL	NULL
3607	142	0.09859155	0.04225352	History	3	NULL
3607	142	0.09859155	0.04225352	Musicology	3	NULL
2203	892	0.8396861	0.03811659	Linguistics	737	NULL
2264	844	0.01658768	0.03436019	NULL	NULL	
1255	2459	0.4221228	0.03416023	Philosophy	954	

parameters:

I've decided to use these parameters now for having only the important elements

```
-- % AH pubs

declare @AH_pubs float = 0.05

-- % discipline / cluster num of pub

declare @Discipline_cluster float = 0.01

-- % discipline / num of AH pub

declare @Discipline_ERIH float = 0.2

drop table #data filter
select *
into #data_filter
from #data
where [Discipline/ERIH] >= @Discipline_ERIH
and [Discipline/Cluster] >= @Discipline_ERIH
and [Score<Perc AH pubs>] >= @AH_pubs
order by [Cluster ID], [Discipline/ERIH]
```

DH pubs

threshold 1%

